

What is claimed is:

- Bele B1*
1. A method for finding documents which relate to a portion of a temporal document, comprising:
    - (a) in response to a signal of interest at a particular time during the temporal document, identifying a portion of the temporal document for which related documents are to be found;
    - (b) selecting text associated with the portion of the temporal document identified;
    - (c) weighting each term in the text selected by a function  $W(t)$  according to the time  $t$  at which the term occurs relative to the time at which the signal of interest occurs;
    - (d) finding the related documents by use of information retrieval techniques as applied to the text selected.
  2. The method of claim 1, wherein the temporal document is video or audio material.
  3. The method of claim 2, wherein the video material is stored on a video server.
  4. The method of claim 2, wherein the text selected is determined by application of speech recognition techniques to the audio component of the portion of the temporal document identified.
  5. The method of claim 2, wherein the text selected is the closed-captioned text associated with the portion of the temporal document identified.
  6. The method of claim 1, wherein the temporal document includes text.
  7. The method of claim 6, wherein the document text appearing to the user varies with time and the text selected is that portion of the temporal document identified.

8. The method of claim 7, wherein the text includes news bulletins, weather, sports scores or stock transaction or pricing information.
9. The method of claim 1, wherein  $W(t)$  is equal for all times between  $t_1$  before the signal of interest is given and  $t_2$  before the signal is given, and is zero for all other times.
10. The method of claim 9, wherein  $t_1$  is 2 seconds and  $t_2$  is 30 seconds.
11. The method of claim 1, wherein  $W(t)$  is equal for all times between  $t_1$  before the signal of interest is given and  $t_2$  before the signal is given, and decreases from  $t_1$  until the time of the signal, and increases from a time  $t_3$  before the signal is given to the time  $t_2$ , and is zero for all other times.
12. The method of claim 11, wherein  $t_1$  is 2 seconds,  $t_2$  is 15 seconds, and  $t_3$  is 30 seconds.
13. The method of claim 11, wherein  $W(t)$  decreases linearly from  $t_1$  until the time of the signal, and increases linearly from  $t_3$  before the signal is given to  $t_2$ .
14. The method of claim 13, wherein  $t_1$  is 2 seconds,  $t_2$  is 15 seconds, and  $t_3$  is 30 seconds.
15. The method of claim 1, wherein
- $$W(t) = W_{t_1, t_2}(t) = \int_{k=0}^t (1 - \exp(-t_1)) * \exp(-t_1 k) * (1 - \exp(-t_2)) * \exp(-t_2(t-k)),$$
- where  $t_1$  and  $t_2$  are constants.
16. The method of claim 15, wherein  $t_1 = .0001$  and  $t_2 = .00025$ .
17. The method of claim 16, wherein  $W(t) = W_{t_1, t_2}(t)$  for times from 30 seconds before the signal of interest is given until the signal is given, and  $W(t) = 0$  for all other times.

18. The method of claim 1, wherein the related documents are accessed through the Internet.

19. The method of claim 18, further including selecting the related documents from among a collection of documents which may be accessed through the Internet, by utilizing databases comprising information about the collection.

20. The method of claim 19, wherein the related documents are selected from the collection according to the scores achieved when evaluating documents in the collection according to a formula giving scores to documents depending upon the occurrence in the documents of terms which occur in text associated with the portion of the temporal document identified, where each term is weighted by a function  $W(t)$  according to the time  $t$  at which the term occurs relative to the time at which the signal of interest occurs.

21. The method of claim 20, wherein a predetermined number of documents, 1000, are selected.

22. The method of claim 20, wherein a score  $S_D$  of a document  $D$  in the collection may be determined by crediting the document  $D$ , for each term  $T$  in the temporal portion of the document identified which occurs in the document  $D$ , with an amount proportional to  $W(t)$ , to Robertson's term frequency  $TF_{TD}$  and to  $IDF_T$  where

$W(t)$  is the weight assigned to a term which occurs at time  $t$  relative to the signal of interest,

$TF_{TD} = N_{TD} / ( N_{TD} + K_1 + K_2 * ( L_D / L_0 ) )$ , and

$N_{TD}$  is the number of times the term  $T$  occurs in document  $D$ ,

$L_D$  is the length of document  $D$ ,

$L_0$  is the average length of a document in the collection of documents indexed,

$K_1$  and  $K_2$  are constants, and

$IDF_T = \log ( (N+K_3) / N_T ) / \log ( N + K_4 )$ , and

$N$  is the number of documents in the collection,

$N_T$  is the number of documents containing the term  $T$  in the collection, and

$K_3$  and  $K_4$  are constants.

23. The method of claim 22, wherein  $K_1$  is 0.5,  $K_2$  is 1.5,  $K_3$  is 0.5, and  $K_4$  is 1.0.

24. The method of claim 20, wherein terms which occur in portions of the temporal document other than the portion identified are utilized in calculating the scores achieved when evaluating documents in the collection.

25. The method of claim 20, wherein the determination of the documents in the collection which receive the highest scores is carried out using compressed document surrogates.

26. The method of claim 16, wherein the determination of the documents in the collection which receive the highest scores is carried out by a server which is distinct from the server which receives the signal of interest.

27. A device for finding documents which relate to a portion of a temporal document, comprising:

(a) means for identifying a portion of the temporal document for which related documents are to be found, in response to a signal of interest at a particular time during the temporal document;

(b) means for selecting text associated with the portion of the temporal document identified;

(c) means for weighting each term in the text selected by a function  $W(t)$  according to the time  $t$  at which the term occurs relative to the time at which the signal of interest occurs;

(d) means for finding the related documents by use of information retrieval techniques as applied to the text selected.

28. The device of claim 27, wherein the temporal document is video or audio material.

29. The device of claim 28, wherein the video material is stored on a video server.
30. The device of claim 28, wherein the text selected is determined by application of speech recognition techniques to the audio component of the portion of the temporal document identified.
31. The device of claim 28, wherein the text selected is the closed-captioned text associated with the portion of the temporal document identified.
32. The device of claim 27, wherein the temporal document includes text.
33. The device of claim 32, wherein the document text appearing to the user varies with time and the text selected is that portion of the temporal document identified.
34. The device of claim 33, wherein the text includes news bulletins, weather, sports scores or stock transaction or pricing information.
35. The device of claim 27, wherein  $W(t)$  is equal for all times between  $t_1$  before the signal of interest is given and  $t_2$  before the signal is given, and is zero for all other times.
36. The device of claim 35, wherein  $t_1$  is 2 seconds and  $t_2$  is 30 seconds.
37. The device of claim 27, wherein  $W(t)$  is equal for all times between  $t_1$  before the signal of interest is given and  $t_2$  before the signal is given, and decreases from  $t_1$  until the time of the signal, and increases from a time  $t_3$  before the signal is given to the time  $t_2$ , and is zero for all other times.
38. The device of claim 37, wherein  $t_1$  is 2 seconds,  $t_2$  is 15 seconds, and  $t_3$  is 30 seconds.

39. The device of claim 37, wherein  $W(t)$  decreases linearly from  $t_1$  until the time of the signal, and increases linearly from  $t_3$  before the signal is given to  $t_2$ .

40. The device of claim 39, wherein  $t_1$  is 2 seconds,  $t_2$  is 15 seconds, and  $t_3$  is 30 seconds.

41. The device of claim 27, wherein

$$W(t) = W_{t_1, t_2}(t) = \int_{k=0}^t (1 - \exp(-t_1)) * \exp(-t_1 k) * (1 - \exp(-t_2)) * \exp(-t_2(t-k)),$$

where  $t_1$  and  $t_2$  are constants.

42. The device of claim 41, wherein  $t_1 = .0001$  and  $t_2 = .00025$ .

43. The device of claim 42, wherein  $W(t) = W_{t_1, t_2}(t)$  for times from 30 seconds before the signal of interest is given until the signal is given, and  $W(t) = 0$  for all other times.

44. The device of claim 27, wherein the related documents are accessed through the Internet.

45. The device of claim 44, further including means for selecting the related documents from among a collection of documents which may be accessed through the Internet, by utilizing databases comprising information about the collection.

46. The device of claim 45, wherein the related documents are selected from the collection according to the scores achieved when evaluating documents in the collection according to a formula giving scores to documents depending upon the occurrence in the documents of terms which occur in text associated with the portion of the temporal document identified, where each term is weighted by a function  $W(t)$  according to the time  $t$  at which the term occurs relative to the time at which the signal of interest occurs.

47. The device of claim 46, wherein a predetermined number of documents, 1000, are selected.

48. The device of claim 46, wherein a score  $S_D$  of a document  $D$  in the collection may be determined by crediting the document  $D$ , for each term  $T$  in the temporal portion of the document identified which occurs in the document  $D$ , with an amount proportional to  $W(t)$ , to Robertson's term frequency  $TF_{TD}$  and to  $IDF_T$  where

$W(t)$  is the weight assigned to a term which occurs at time  $t$  relative to the signal of interest,

$TF_{TD} = N_{TD} / ( N_{TD} + K_1 + K_2 * ( L_D / L_0 ) )$ , and

$N_{TD}$  is the number of times the term  $T$  occurs in document  $D$ ,

$L_D$  is the length of document  $D$ ,

$L_0$  is the average length of a document in the collection of documents indexed,

$K_1$  and  $K_2$  are constants, and

$IDF_T = \log ( (N + K_3) / N_T ) / \log ( N + K_4 )$ , and

$N$  is the number of documents in the collection,

$N_T$  is the number of documents containing the term  $T$  in the collection, and

$K_3$  and  $K_4$  are constants.

49. The device of claim 48, wherein  $K_1$  is 0.5,  $K_2$  is 1.5,  $K_3$  is 0.5, and  $K_4$  is 1.0.

50. The device of claim 46, wherein terms which occur in portions of the temporal document other than the portion identified are utilized in calculating the scores achieved when evaluating documents in the collection.

51. The device of claim 46, wherein the determination of the documents in the collection which receive the highest scores is carried out using compressed document surrogates.

Express Mail No EK220865491US

GTE Ref. No. 99-849

52. ~~The device of claim 42, wherein the determination of the documents in the collection which receive the highest scores is carried out by a server which is distinct from the server which receives the signal of interest.~~